Next Stop Glycomics

In the postgenomic era, the world of protein glycosylation may be the next place to go.

BY MARK S. LESNEY

In the distance, a branched and spiky forest of carbohydrate thorns arises from a host of proteins and the vast majority of cells and organs. Sugar "coatings" create a landscape of recognition sites, barriers, and carriers that help control the rhythms of metabolism from conception to catabolism. Carbohydrates control immunological recognition, pathogen attack, and even protein folding and placement (*1*).

On the pharmaceutical front, the need to navigate through the carbohydrate jungle may be one reason that so many drug candidates fail to meet expectations, and why so many diseases have yet to reveal a viable target for a useful therapeutic agent. Indeed, this jungle may conceal the ultimate secret to pharmacogenomics—why individuals respond so differently to the same drug compounds.

All this has led researchers to declare that the next stop on the 'omics trail should be glycomics—the study of the complete carbohydrate complement of an organism. Why now instead of earlier? One key reason is a change from the excessive optimism embedded in proteomics and in the simplistic view of the nature and role of "bare" proteins. (Ironically, this is the same attitudinal problem that required people to abandon an excessive trust in the benefits of genomics and put them on the proteomics trail.)

Glycoproteins

Polysaccharides and glycolipids are critical to cell function and fall under the rubric of glycomics. But arguably, proteins with glycans attached are the most interesting of the carbohydrate-containing moieties in a cell as well as some of the most difficult to study.

Glycans are typically attached to proteins in one of three fashions. They are normally either N-linked to asparagine (Asn) residues in the peptide chain occurring in the sequence Asn-X-Thr/Ser, or O-linked to serine (Ser) or threonine (Thr) residues;

ILLUSTRATION: GEOFF SMITH

© 2002 AMERICAN CHEMICAL SOCIETY

Advertisers

BIO-RAD LABORATORIES
1000 Alfred Nobel Drive, Hercules, CA 94547, 800-424-6723, www.biorad.com

AMERSHAM BIOCEIENCES
Amersham Place, Little Chalfont, Buckinghamshire HP7 9NA England
0870 606 1921, 01494 544350
www.amershambiosciences.com

COMPUGEN LTD.
72 Pinchas Rosen Street, Tel Aviv, Israel 69512, 972-3-7658-585, 972-3-7658-555, www.compugen.co.il

alternatively, they are attached to the terminal carboxyl of the peptide chain of some membrane proteins. Glycosylation in cells occurs in the endoplasmic reticulum and the Golgi apparatus.

N-linked glycans originate from the attachment of a conserved 14-sugar scaffold that acts as the basis for subsequent addition and removal of other sugars. Two forms of N-linked glycoproteins are produced: those attached to a high-mannose oligosaccharide (which consists of 10 of the 14 original attached sugars), and those with complex oligosaccharides that are formed via significant modifications of the original scaffold. O-linked glycosylation has no specific precursor, although an N-acetylglucosamine is generally attached first to a chain that is variable in number and type of sugars (generally from a few to 10 or more).

In both N-linked complex oligosaccharides and the O-linked glycoproteins, the addition of each subsequent sugar massively increases the number of possible conformational states, as attachment can theoretically occur via any of the six available hexose carbons. The genetic and biochemical control of carbohydrate attachment is so complex that not only are there more than 200 putative glycotransferase sequences in GenBank, but for each possible linkage and for each type of sugar, "there is at least one and often a family of glycotransferases that creates that structure," says Linda Baum, a researcher at the University of California at Los Angeles (*1*). To date, 110 genes that encode glycotransferases and related genes have been cloned, and in some cases they have been used to create knockout (gene-deficient) mice and cell lines for functional analysis (*2*).

Sugar-coated secrets

Despite years of research, much of the role of glycans in cellular physiology, cell–cell interactions, reproduction, and disease is still shrouded in mystery. Some researchers speculate that the manipulation of sugars (and the recognition and communication this allows) has led to the complexity possible in multicellular life forms. In this vein, posttranslational glycoprocessing may even account for part of the versatility of what appears to be a surprisingly low number of genes in the human genome.

Appropriate sugar attachment and processing appear critical to correct protein folding as well as to the ultimate localization of proteins. And attached sugars are known to protect certain proteins from a variety of proteases—indicating a role in controlling protein turnover.

Sweet poisons and palliatives

The role of glycans in disease is a complicated one. According to William Hancock (*3*), editor of the *Journal of Proteome Research,* "It is difficult to determine whether one is observing cause or effect, but there is no question that changes in glycosylation are strongly associated with many different forms of cancer, as well as other diseases such as rheumatoid arthritis." For example, there is often a differing degree of fucosylation between cancer and noncancer cells. Surface sugar differences between cancerous and healthy cells have even demonstrated pharmacological significance. Certain breast and ovarian cancer cells produce an overabundance of the glycosylated form of ceramide (glycosylceramide). Because of this glycosylation, these cancer cells escape ceramide-induced cell death, making them resistant to the chemotherapeutic drug Adriamycin (*4*).

Diabetes provides an example of a disease in which excess glycosylation appears responsible for some of the symptomatic damage. Nonenzymatic glycosylation, made possible by the accumulation of excess sugars, has been found to damage key lens proteins, contributing to the development of cataracts. A large number of diseases also are linked to errors in the metabolism of glycosylated compounds; many such errors are due to the accumulation of glycosylated byproducts. These maladies include Fabry's and Gaucher's diseases.

In infectious diseases, viral and bacterial pathogens recognize specific glycosylated sites on cell surfaces. The drugs Relenza and Tamiflu act to limit the spread of the flu in humans by inhibiting a viral neuraminidase, which normally allows newly formed influenza virus particles to bud free of these surface binding sites to infect other cells.

On the drug production side, it is difficult, because of microheterogeneity, to detect all the glycan types attached to the same individual proteins (which can exist as a variably glycosylated "family"). These subtle differences may not be detectable by lectins or even monoclonal antibodies. And the individual "family members" may occur in such small amounts as to be undetectable by standard analytical methods. Recombinant DNA-derived protein drugs, such as tissue plasminogen activator, are glycosylated, and the FDA requires strict monitoring of batches to ensure production of the approved therapeutic forms. Inadequate glycosylation forced Amgen to discard up to 80% of its recombinant erythropoietin until the company found a means to add two sugar residues to those normally found to create the longer-acting version sold as Aranesp (*5*).

The first major differentiation in glycoprotein function depends on whether the glycans are N- or O-linked. One tantalizing clue to their function involves the realization that O-linked glycans are attached at the same sites where protein phosphorylation also typically occurs. As protein phosphorylation is critical to the regulation of innumerable metabolic pathways, there might be a possible regulatory role for glycosylation as well.

Although, in many cases, knockout cell lines appear to have little problem with major disruptions in glycoprotein production, the interactions of cells that are necessary in whole organs and organisms seem to be disrupted. This may explain the increasingly apparent role of glycans in disease (see box, "Sweet poisons and palliatives").

Enzymes and instruments

Perhaps the most daunting barrier on the road to glycomics has been a relative lack of available techniques and instrumentation to perform rapid and reliable analyses. Because of the many permutations (including branching) possible for polysaccharide linkages compared with those in peptides and nucleic acids, glycan composition and even sequencing do not yield significant information on 3-D structure. This is a significant problem because the 3-D configuration of the glycosylated portion of the glycoprotein can be responsible for most of its unique properties.

Lectins and enzymes. One of the earliest discovered means of

studying glycoproteins was to characterize them according to their unique interactions with a variety of specific, naturally occurring, glycan-binding proteins known as lectins. Each lectin recognizes a unique sugar pattern in an antibody-like manner. Plant lectins, the earliest to be discovered, rapidly became a major tool for diagnostic identification of particular glycan structures. Today, the use of mammalian lectins has greatly expanded the repertoire of these molecules in glycoprotein analysis.

More detailed analysis became possible with the discovery of the multitude of glycosidases, many of which had unique specificities. Exoglycosidases, for example, often cut only one particular terminal sugar attached in a specific way to the glycan chain, and many of the endoglycosidases only cut between specific sugar sequences, which is similar to the behavior of a DNA restriction enzyme. This allows for glycan sequencing and, in some instances, for the reconstruction of 3-D structure based on subsequent chromatographic or MS analysis of enzymatic digests.

As interest in glycoproteins and equipment sophistication grew, so did the use of tryptic digests to break proteins into peptide fragments. Chromatography or 2-D electrophoresis to isolate the glycosylated fragments allowed for chemical analysis of the 3-D structure of the sequestered glycans and their attached peptides by NMR spectroscopy or MS (*5*). The conservation of many glycan structures allowed for libraries of MS fragmentation patterns to be developed that could give clues for identifying known glycoproteins—but these were generally insufficiently robust to determine the 3-D structures of complex unknowns. Analysis of such unknowns generally became a full-scale research project, requiring the gamut of technologies used to analyze chemical structures, massive amounts of labor, and significant computational analysis just to get started.

The need for the move to glycomics shows the limitations of gene-only approaches to physiology research and drug discovery.

Mass spectrometry. With such difficulties, glycobiology (as it was named in 1988) seems doomed to never become glycomics, to never achieve the power of proteomics or genomics until the critical problems of automated synthesis and structural analysis are solved. There are indications, though, that at least a start toward such solutions is at hand. The development of a variety of multiple-tandem MS (MS^n) techniques may have already dented, if not broken, the structural determination barrier.

Researchers at ThermoFinnigan reported on developing a high-throughput system for structural analysis of glycosylated proteins using multiple-tandem ion-trap MS (*6*). In their method, the product ion spectra from a tryptic digest of tissue plasminogen activator were analyzed using TurboSEQUEST software (see Sites and Software, p 21). Putative glycopeptides were subjected to further breakdown through MS^4, allowing analytical fragmentation for identification of carbohydrate residues. These could be matched to a database for analysis. The significant benefits of such an analytical system are of course predicated on the quality and comprehensiveness of the databases used for comparison identification. This points to a key problem in glycomics—such databases are not currently available for the tremendous variety of molecular conformations possible. According to William Hancock, one of the authors of this report, a variety of approaches will also likely be necessary to move forward in studying glycosylation. These include 2-D analysis, MS systems, microarrays, and the use of lectins and glycosidases, as well as the development of extensive bioinformatics capabilities (*3*).

New assays and new chemistries

On the synthesis side of the glycomics problem, research into glycan construction using combinatorial chemistry is proceeding apace, based on work begun more than 25 years ago. Both standard chemical synthesis and enzymatic techniques, sequentially using the manifold glycotransferases, have been used routinely (*7*). One of the most recent developments has been to investigate the use of glycosamino acids as building blocks for solid-phase combinatorial synthesis of new drug candidates (*8*). Still, profound limitations exist in the exact duplication of many complex 3-D structures because of the problem of controlling the inherent flexibility of oligosaccharide linkages.

But pharmaceutical progress cannot and has not waited for a complete solution. Not only have a variety of glycoconjugate drugs already been discovered—and improved—but the development of the recombinant protein drugs that require effective glycosylation has forced the issue of understanding and controlling glycosylation. In addition, new lectins are being searched for, and new means for screening glycan targets are being deployed—specifically, glycochip arrays, as developed by Glycominds Ltd. (*www.glycominds.com*).

The need for the move to glycomics shows the limitations of gene-only approaches to physiology research and drug discovery. Posttranslational processing of proteins is demonstrably ever more important; the complex metabolic, structural, and regulatory aspects of the living cell that involve glycans demonstrate that the work of biomedicine lies fully ahead of us—genomics being just the barest start. With regard to long-term employment prospects, however, that may turn out to be a sweet deal for researchers after all.

References

(1) Baum, L. G. *Immunity* **2002,** *16,* 5–8.
(2) Taniguchi, N. Beyond Glycogenes; www.glycoforum.gr.jp/science/glycogenes/BG01/BG01E.html.
(3) Hancock, W. S. *J. Proteome Res.* **2002,** *1,* 297.
(4) Liu, Y.-Y.; Cabot, M. C. Reversal of Adriamycin Resistance. Presented at the California Breast Cancer Symposium, Oakland, CA, March 8–10, 2002; www.ucop.edu/srphome/bcrp/symposium/2002/abstracts/a/a_13.html.
(5) Snyder, A. P. *Interpreting Protein Mass Spectra;* American Chemical Society: Washington, DC, 2000.
(6) Wu, S.-L.; et al. *Structural analysis of glycosylated peptides in complex mixtures with ion trap MS^n;* ThermoFinnigan Application Report No. 300; available at www.thermofinnigan.com.
(7) *Combinatorial Chemistry and Molecular Diversity in Drug Discovery;* Gordon, E. M., Kerwin, J. F., Eds.; Wiley-Liss: New York, 1998.
(8) Schweitzer, F. *Angew. Chem., Int. Ed.* **2002,** *41,* 230–253.

Mark S. Lesney is a senior associate editor of *Modern Drug Discovery*. Send your comments or questions about this article to mdd@acs.org or to the Editorial Office address on page 3. ■